

Universal Multiple-Octet Coded Character Set  
 International Organization for Standardization  
 Organisation internationale de normalisation  
 Международная организация по стандартизации

**Doc Type: Working Group Document****Title: Proposal to encode the Avestan script in the BMP of the UCS****Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)****Authors: Michael Everson and Roozbeh Pournader****Status: Liaison Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Replaces: N2566, N1684****Date: 2006-10-20**

**1. Introduction.** The Avestan script was used to write two kinds of text: religious texts from the *Avesta*, written in the Avestan language (in two variants, Old Avestan and Young Avestan), and texts in Pāzand (that is, Middle Persian written in the Avestan script). The script was derived in the fifth or sixth century CE (some writers say as early as the fourth century CE) as a rationalization and improvement on the original (Book) Pahlavi script, which itself was derived from a variety of Aramaic. The Zoroastrians used modified letterforms to distinguish between signs which in Pahlavi had fallen together. Although Pahlavi script is used alongside Avestan in text, it behaves differently and the two cannot be unified. Pahlavi has cursive joining behaviour; Avestan does not. In Avestan, sometimes letters touch in kerning, but more often they do not. There are a few ligatures in Avestan, discussed below. Pahlavi is not further considered in this proposal.

**2. Structure.** Avestan is a simple alphabetic script written from right to left. Avestan letters are written separately, or touch in close kerning. Four ligatures are commonly used in manuscripts of the *Avesta*: 𐬀𐬀 SHE + 𐬀 A = 𐬀𐬀 ša (unligated 𐬀𐬀); 𐬀 SHE + 𐬀 CE = 𐬀𐬀 šc (unligated 𐬀𐬀); 𐬀 SHE + 𐬀 TE = 𐬀𐬀 št (unligated 𐬀𐬀); 𐬀 A + 𐬀 HE = 𐬀𐬀 ah (unligated 𐬀𐬀). Fonts which implement ligatures will typically treat these combinations as default ligatures, which can be broken with ZWNJ. The choice of representation of these ligatures in printed texts is a matter for the editor.

**3. Character set.** The Avestan alphabet is well-attested, and is extended here by only one character. The letter 𐬀 RE is modified to 𐬀 LE (or 𐬀 as in Figure 3) for use in Pāzand to represent [l] (see Figures 6 and 11). Some writers consider the digraphs 𐬀 ii and 𐬀 uu as letters and transliterate them as y and v, but they are properly encoded as two-character strings.

**4. Punctuation.** Avestan punctuation displays the usual inconsistency of any early writing system. The punctuation proposed here for encoding is based on the system Geldner established in his 1880 edition of the *Avesta*. The 𐬀 AVESTAN SEPARATION POINT follows each word in Avestan and is usually followed by a space (as is most Avestan punctuation); it is *not* a sentence delimiter like U+002E FULL STOP, and indeed Geldner uses it alongside the FULL STOP in his edition (see Figure 15). The 𐬀 AVESTAN COLON, 𐬀 AVESTAN SEMICOLON, and 𐬀 AVESTAN END OF SENTENCE indicate breaks of increasing finality, followed by 𐬀 AVESTAN END OF SECTION which itself may be doubled 𐬀𐬀 (sometimes with a space between 𐬀 𐬀) for extra finality (see Figures 16 and 17). The 𐬀 AVESTAN TURNED END OF SENTENCE and 𐬀 AVESTAN TURNED END OF SECTION are also attested, but were not used by Geldner. Sometimes the AVESTAN SEPARATION POINT may precede, or follow, other punctuation. Geldner says that the 𐬀 AVESTAN ABBREVIATION MARK “is employed for a sign

of abbreviation and repetition” and indeed he uses it extensively in his footnotes “for the sake of economizing space”.

**5. Linebreaking.** Geldner 1880 shows no hyphenation. In the manuscripts words may be broken arbitrarily, but it is unlikely that modern users will prefer this behaviour. A space normally follows most Avestan punctuation, and one may certainly expect a linebreak in such a position.

**6. Ordering.** No ordering is attested in antiquity, but the codepoint order of the Avestan letters reflects the most common current Iranianist practice (see Figure 2), which may have been established by the Parsee community itself. This suffices for all of the letters except 𐬶 LE, which should order on the second level with 𐬷 RE.

**7. Character names.** Names here are based on the transliteration values and are intended to be mnemonic.

**8. Unicode Character Properties.** Character properties are proposed here.

```
0800;AVESTAN LETTER A;Lo;0;R;;;;N;;;;;
0801;AVESTAN LETTER AA;Lo;0;R;;;;N;;;;;
0802;AVESTAN LETTER AO;Lo;0;R;;;;N;;;;;
0803;AVESTAN LETTER AAO;Lo;0;R;;;;N;;;;;
0804;AVESTAN LETTER AN;Lo;0;R;;;;N;;;;;
0805;AVESTAN LETTER AAN;Lo;0;R;;;;N;;;;;
0806;AVESTAN LETTER AE;Lo;0;R;;;;N;;;;;
0807;AVESTAN LETTER AEE;Lo;0;R;;;;N;;;;;
0808;AVESTAN LETTER E;Lo;0;R;;;;N;;;;;
0809;AVESTAN LETTER EE;Lo;0;R;;;;N;;;;;
080A;AVESTAN LETTER O;Lo;0;R;;;;N;;;;;
080B;AVESTAN LETTER OO;Lo;0;R;;;;N;;;;;
080C;AVESTAN LETTER I;Lo;0;R;;;;N;;;;;
080D;AVESTAN LETTER II;Lo;0;R;;;;N;;;;;
080E;AVESTAN LETTER U;Lo;0;R;;;;N;;;;;
080F;AVESTAN LETTER UU;Lo;0;R;;;;N;;;;;
0810;AVESTAN LETTER KE;Lo;0;R;;;;N;;;;;
0811;AVESTAN LETTER XE;Lo;0;R;;;;N;;;;;
0812;AVESTAN LETTER XYE;Lo;0;R;;;;N;;;;;
0813;AVESTAN LETTER XVE;Lo;0;R;;;;N;;;;;
0814;AVESTAN LETTER GE;Lo;0;R;;;;N;;;;;
0815;AVESTAN LETTER GGE;Lo;0;R;;;;N;;;;;
0816;AVESTAN LETTER GHE;Lo;0;R;;;;N;;;;;
0817;AVESTAN LETTER CE;Lo;0;R;;;;N;;;;;
0818;AVESTAN LETTER JE;Lo;0;R;;;;N;;;;;
0819;AVESTAN LETTER TE;Lo;0;R;;;;N;;;;;
081A;AVESTAN LETTER THE;Lo;0;R;;;;N;;;;;
081B;AVESTAN LETTER DE;Lo;0;R;;;;N;;;;;
081C;AVESTAN LETTER DHE;Lo;0;R;;;;N;;;;;
081D;AVESTAN LETTER TTE;Lo;0;R;;;;N;;;;;
081E;AVESTAN LETTER PE;Lo;0;R;;;;N;;;;;
081F;AVESTAN LETTER FE;Lo;0;R;;;;N;;;;;
0820;AVESTAN LETTER BE;Lo;0;R;;;;N;;;;;
0821;AVESTAN LETTER BHE;Lo;0;R;;;;N;;;;;
0822;AVESTAN LETTER NGE;Lo;0;R;;;;N;;;;;
0823;AVESTAN LETTER NGYE;Lo;0;R;;;;N;;;;;
0824;AVESTAN LETTER NGVE;Lo;0;R;;;;N;;;;;
0825;AVESTAN LETTER NE;Lo;0;R;;;;N;;;;;
0826;AVESTAN LETTER NYE;Lo;0;R;;;;N;;;;;
0827;AVESTAN LETTER NNE;Lo;0;R;;;;N;;;;;
0828;AVESTAN LETTER ME;Lo;0;R;;;;N;;;;;
0829;AVESTAN LETTER HME;Lo;0;R;;;;N;;;;;
082A;AVESTAN LETTER YYE;Lo;0;R;;;;N;;;;;
082B;AVESTAN LETTER YE;Lo;0;R;;;;N;;;;;
082C;AVESTAN LETTER VE;Lo;0;R;;;;N;;;;;
082D;AVESTAN LETTER RE;Lo;0;R;;;;N;;;;;
082E;AVESTAN LETTER LE;Lo;0;R;;;;N;;;;;
082F;AVESTAN LETTER SE;Lo;0;R;;;;N;;;;;
0830;AVESTAN LETTER ZE;Lo;0;R;;;;N;;;;;
0831;AVESTAN LETTER SHE;Lo;0;R;;;;N;;;;;
0832;AVESTAN LETTER ZHE;Lo;0;R;;;;N;;;;;
0833;AVESTAN LETTER SHYE;Lo;0;R;;;;N;;;;;
0834;AVESTAN LETTER SSHE;Lo;0;R;;;;N;;;;;
0835;AVESTAN LETTER HE;Lo;0;R;;;;N;;;;;
0838;AVESTAN ABBREVIATION MARK;Po;0;R;;;;N;;;;;
0839;AVESTAN SEPARATION POINT;Po;0;R;;;;N;;;;;
083A;AVESTAN COLON;Po;0;R;;;;N;;;;;
083B;AVESTAN SEMICOLON;Po;0;R;;;;N;;;;;
```

083C;AVESTAN END OF SENTENCE;Po;0;R;;;;;N;;;;;  
 083D;AVESTAN TURNED END OF SENTENCE;Po;0;R;;;;;N;;;;;  
 083E;AVESTAN END OF SECTION;Po;0;R;;;;;N;;;;;  
 083F;AVESTAN TURNED END OF SECTION;Po;0;R;;;;;N;;;;;

## 9. Bibliography.

- Abolghassemi, Mohsen. 1995 (1374 AP). *Riše-Šenāsi (Etymology)*. Tehran: Qoqnoos. ISBN 964-311-045-1.
- Adib-Solṭāni, M. Š. 2000 (1378 AP). *An Introduction to Problems of Persian Orthography*. Third edition. Tehrān: Amir Kabir. ISBN 964-00-0611-4.
- Fossey, Charles. 1948. *Notices sur les caractères étrangers anciens et modernes rédigées par une groupe de savants*. Nouvelle édition mise à jour à l'occasion du 21e Congrès des Orientalistes. Paris: Imprimerie Nationale de France.
- Geldner, Karl F. 1880. *Avesta: the sacred books of the Parsis*. Stuttgart: W. Kohlhammer. Reprinted in 2003 with an introduction in Persian by Dr Jaleh Amouzgar.
- Haarmann, Harald. 1990. *Universalgeschichte der Schrift*. Frankfurt/Main; New York: Campus. ISBN 3-593-34346-0.
- Hoffmann, Karl, and B. Forssman. 1996. *Avestische Laut- und Flexionslehre*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Kōno Rokurō, Chino Eiichi, & Nishida Tatsuo. 2001. *The Sanseido Encyclopaedia of Linguistics. Volume 7: Scripts and Writing Systems of the World [Gengogaku dai ziten (bekkan) sekai mozi ziten]*. Tokyo: Sanseido Press. ISBN 4-385-15177-6.
- Oryan, Said. 1999 (1377 AP). *Pahlavi-Pazand glossary: Farhang ī Pahlavi*. (Language and Literature; 4) Tehran: Research Institute for Islamic Culture and Art. ISBN 964-471-414-8.
- Pietraszewski, Ignacy. 1858–1862. *Zend-avesta ou plutôt Zen-daschta expliqué d'après un principe tout à fait nouveau par I.* Berlin: Chez l'Auteur Éditeur.
- Poure Davoud, Ebrahim. 1927. *Gāthās, the Oldest Parts of the Avesta*. Bombay: Bombay's Zoroastrian Society of Iranians. Reprinted in 1998 (1377 AP) Tehran: Asatir. ISBN 964-5960-93-2.
- Rashed Mohassel, Mohammad Taghi. 2003 (1382 AP). *The Avesta: Praise to Truth and Purity*. Tehran: Cultural Research Bureau. ISBN 964-379-008-8.
- Reichelt, Hans. 2004 (1383 AP). *Avesta reader: an approach to the Zoroaster's Gathas and new Avestan texts*. Translated and annotated with Persian translation of hymns and texts by Jalil Doostkhah. Tehran: Qoqnoos Publishing. ISBN 964-311-473-2.
- Reichsdruckerei. 1924. *Alphabete und Schriftzeichen des Morgen- und Abendlandes, zum allgemeinen Gebrauch mit besonderer Berücksichtigung des Buchgewerbes*. Unter Mitwirkung von Fachgelehrten zusammengestellt in der Reichsdruckerei. Berlin: Reichsdruckerei.

## Acknowledgements

This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley), as well as by support from the Science and Arts Foundation, Iran, and from the High Council of Information Dissemination of Iran (which funded the FarsiWeb Project, Sharif FarsiWeb, Tehran).

## Figures

1	𐬀	a	17	𐬁	k	35	𐬌	n
2	𐬁	ā	18	𐬂	x	36	𐬍	n̄
3	𐬂	ā̇	19	𐬃	ḫ	37	𐬎	n̄ <sup>v</sup>
4	𐬃	ā̇̇	20	𐬄	x <sup>v</sup>	38	𐬏	n
5	𐬄	q	21	𐬅	g	39	𐬐/𐬑	n̄
6	𐬅/𐬆	q̇	22	𐬆	ġ	40	𐬇	n̄
7	𐬇	o	23	𐬈	r	41	𐬈	m
8	𐬈	ō	24	𐬉	c	42	𐬉	m̄
9	𐬉	e	25	𐬊	j	43	𐬊	y
10	𐬊	ē	26	𐬋	t	44	𐬋	v
11	𐬋	o	27	𐬌	θ	45	𐬌	r
12	𐬌	ō	28	𐬍	d	46	𐬍	š
13	𐬍	i	29	𐬎/𐬏	ḏ	47	𐬎	s
14	𐬏	ī	30	𐬐/𐬑	t̄	48	𐬐	z
15	𐬑	u	31	𐬑	p	49	𐬑	š
16	𐬒	ū	32	𐬒	f	50	𐬒	ž
			33	𐬓	b	51	𐬓	š
			34	𐬔	β	52	𐬔	y
						53	𐬕(𐬖)	h

Figure 1. Avestan alphabet from Geldner's 1880 edition of the *Avesta*.

## Schrifttafel der Avesta-Schrift

1 𐬀 <i>a</i>	2 𐬁 <i>ā</i>	3 𐬂 <i>ā̇</i>	4 𐬃 <i>ā̈</i>	5 𐬄 <i>ȧ</i>	6 𐬅 𐬆 <i>ā̇</i>	7 𐬇 <i>ə</i>	8 𐬈 <i>ē</i>		
		9 𐬉 <i>e</i>	10 𐬊 <i>ē</i>	11 𐬋 <i>o</i>	12 𐬌 <i>ō</i>	13 𐬍 <i>i</i>	14 𐬎 <i>ī</i>	15 𐬏 <i>u</i>	16 𐬐 <i>ū</i>
17 𐬑 <i>k</i>	18 𐬒 <i>x</i>	19 𐬓 <i>ẋ</i>	20 𐬔 <i>x<sup>v</sup></i>	21 𐬕 <i>g</i>	22 𐬖 <i>ġ</i>	23 𐬗 <i>γ</i>			
24 𐬘 <i>c</i>	25 𐬙 <i>j</i>								
26 𐬚 <i>t</i>	27 𐬛 <i>θ</i>	28 𐬜 <i>d</i>	29 𐬝 𐬞 <i>δ</i>	30 𐬟 𐬠 <i>ṫ</i>					
31 𐬡 <i>p</i>	32 𐬢 <i>f</i>	33 𐬣 <i>b</i>	34 𐬤 <i>β</i>						
35 𐬦 <i>ŋ</i>	36 𐬧 <i>ŋ̇</i>	37 𐬨 <i>ŋ<sup>v</sup></i>	38 𐬩 <i>n</i>	39 𐬪 <i>ṅ</i>	40 𐬫 <i>n̈</i>	41 𐬬 <i>m</i>	42 𐬭 <i>ṁ</i>		
43 𐬮 <i>ẏ</i>	44 𐬯 <i>y</i>	45 𐬰 𐬱 <i>v</i>	46 𐬲 <i>r</i>						
47 𐬳 <i>s</i>	48 𐬴 <i>z</i>	49 𐬵 <i>š</i>	50 𐬶 <i>ž</i>	51 𐬷 <i>ṧ</i>	52 𐬸 <i>ž̇</i>				
53 𐬹 𐬺 <i>h</i>									

**Figure 2.** Avestan alphabet according to Hoffmann & Forssman 1996.

gl [gil]                      گل، لجن                      ګل، لجن

[ g ǣ. gīl. gīll. gīlln ] gilln : پازند: gil، پازند: [ g ǣ. gīl. gīll. gīlln ]

pp                      pd                      f

Hau. 219, Māck. 36, Junk. 119

**Figure 3a.** Example in the second line (the first is Pahlavi) of doubled LETTER LE in ګلن *gilln* from Oryan 1999. The double bar does not appear to have significance. Oryan ligates in his handwriting, and uses the stroke form ʃ rather than the ring form ʃ̌.

hlws/xraōs                      نکی                      ځلوس/خراوس

LKA / llnaka                      نکی                      ګلوس

**Figure 3b.** Example in the fourth line of doubled LETTER LE in ګلن *llnakā* from Oryan 1999.

LKA[būm]                      بوم، سرزمین، قلمرو خاکی، ولایت                      ګلوس

آرامی باستان، آرامی امپراطوری، ماندایی: arqā، صورت کهن تر:

arā، نبطی، پالمیری: rē، آرامی کتاب مقدس: ʿēraṣ=ʿāraq سرياني:

rē، پازند: [lakā] llnakā

f                      f

Gese.1083, Nyb.41, 62, Junk.107

**Figure 3c.** Example in the second line of doubled LETTER RE used for [l] in ګلن *rrnakā* from Oryan 1999.

MYLA -k[namak]                      نمک                      ګلوس

عبری: melah = عبری نو، آرامی، سرياني: milḥā، آرامی باستان، آرامی

امپراطوری، پالمیری، آرامی کتاب مقدس، آرامی یهودی، فلسطینی

مسیحی، ماندایی: milḥā برای milā > milḥā، مقایسه شود با ماندایی

mālai mailāy، پازند: ملح، عربی: miṣḥā > miṣā

[mailāi. m(a)illnāy

f                      pp                      s                      m

Nyb. 48, 87, Hau. 153, 159, Gese. 571, Sm. 276, Dro. 266, 270,

f

Junk.115

MYLA- k/ mailāy                      نکی                      ګلوس

MYLA-k/mailāy                      نکی                      ګلوس

**Figure 3d.** Example in the second line of a single LETTER LE in ګلن *mailāy* from Oryan 1999. Compare ګلن *mairāy* in the fourth line and ګلن *millnāy* in the sixth.



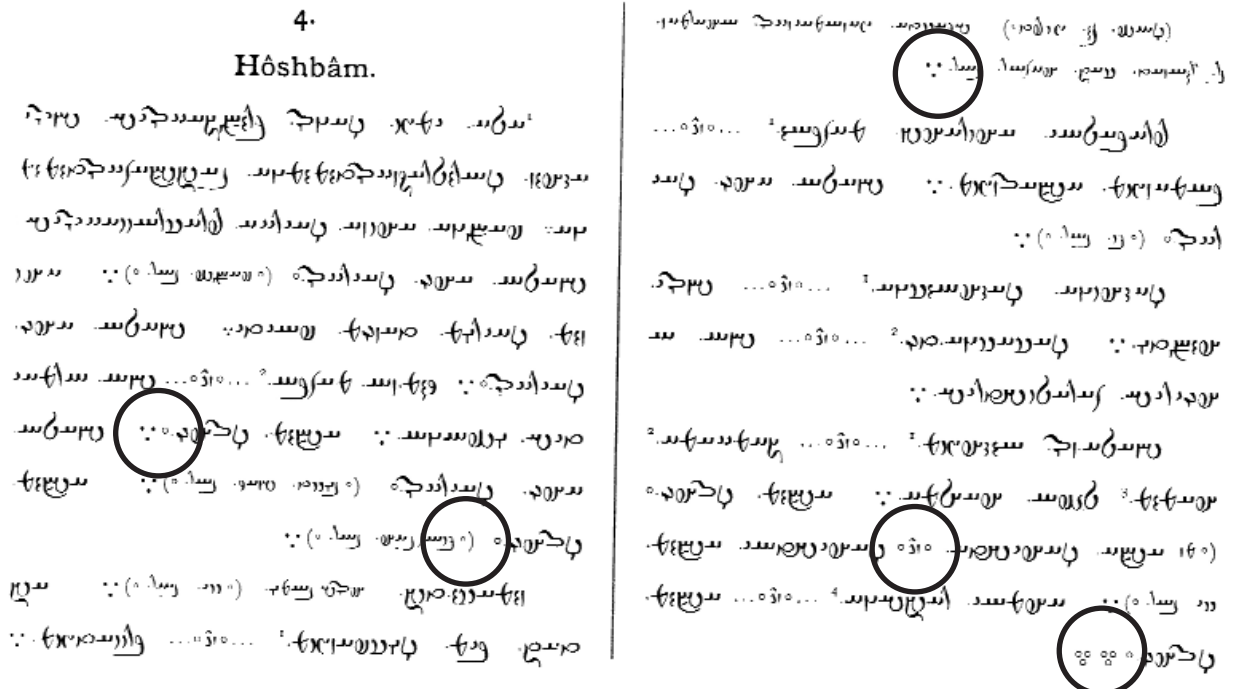


Figure 4. Avestan text from Geldner 1880, showing examples of punctuation marks.

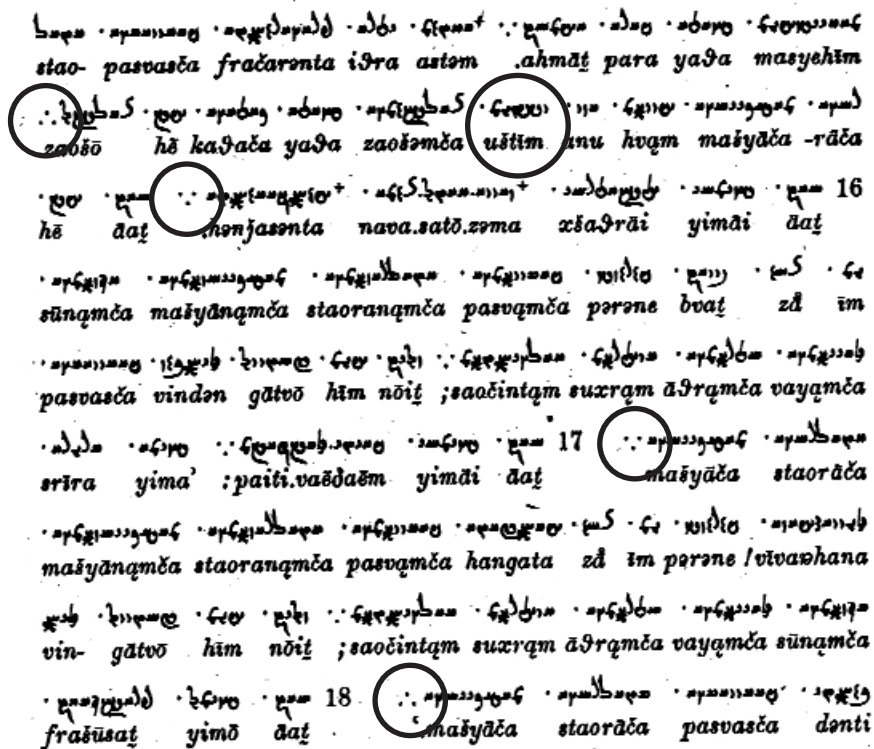


Figure 5. Avestan text from Reichelt 2004, showing two kinds of punctuation.

Also shown is the *št* ligature 𐬰𐬀 in 𐬰𐬀𐬀𐬀 *uštīm*. Reichelt does not use the ligature 𐬰𐬀 *ah*.

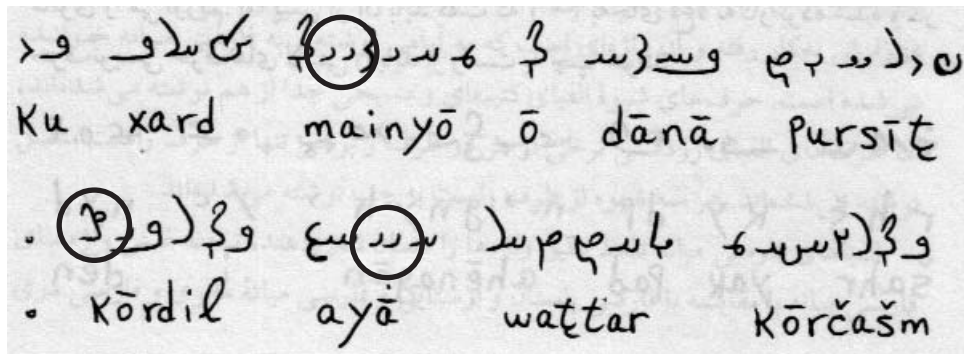


Figure 6. Handwritten sample of Avestan text with transliteration from Abolghassemi 1995, showing 𐬰 for *l* as well as 𐬀 *ii* transliterated as *y*.

### جدول III. الفباى اوستايى يا «دين ديڤرى»\*

#### حرفهاى الفبا

1 𐬀	2 𐬁	3 𐬂	4 𐬃	5 𐬄	6 𐬅	7 𐬆	8 𐬇	9 𐬈
10 𐬉	11 𐬊	12 𐬋	13 𐬌	14 𐬍	15 𐬎	16 𐬏	17 𐬐	18 𐬑
19 𐬒	20 𐬓	21 𐬔	22 𐬕	23 𐬖	24 𐬗	25 𐬘	26 𐬙	27 𐬚
28 𐬛	29 𐬜	30 𐬝	31 𐬞	32 𐬟	33 𐬠	34 𐬡	35 𐬢	36 𐬣
37 𐬤	38 𐬥	39 𐬦	40 𐬧	41 𐬨	42 𐬩	43 𐬪	44 𐬫	45 𐬬
46 𐬭	47 𐬮	48 𐬯	49 𐬰	نویسه‌هاى پیوسته				

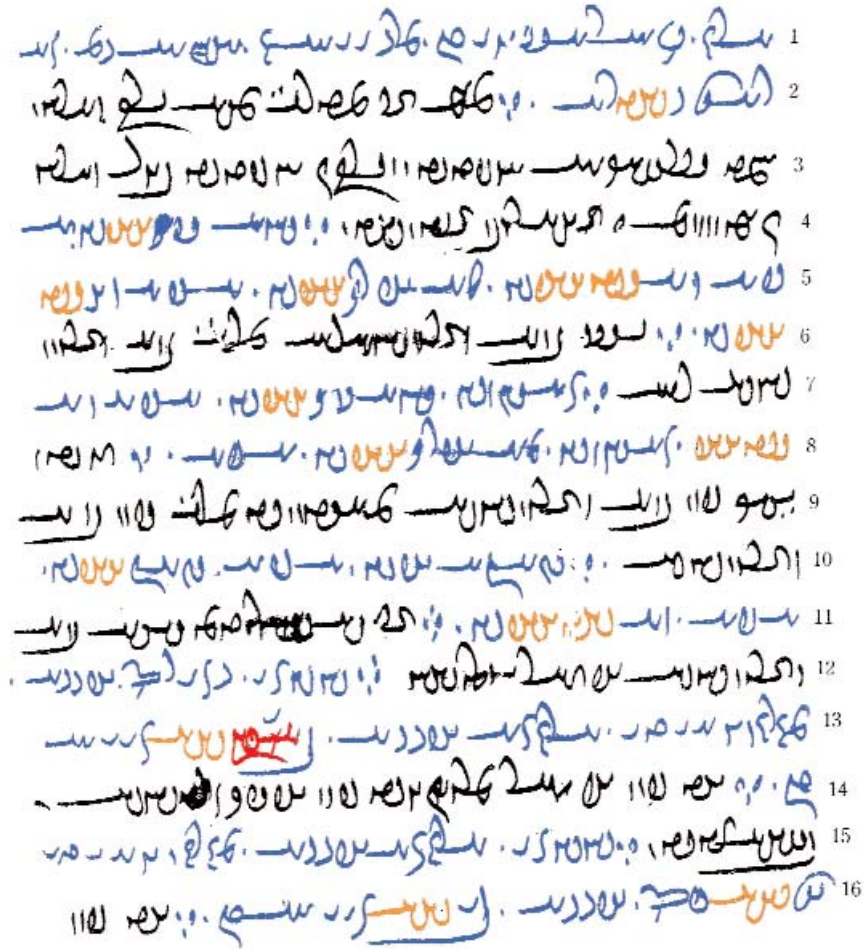
Figure 7. Avestan alphabet from Adib-Soltani 2000, showing 𐬥 *ii* and 𐬬 *uu*. Also shown are the ligatures *ša* 𐬱, *šc* 𐬲, and *št* 𐬳.

#### (210) Das Awesta-Alphabet

Zeichen	Lautwert	Zeichen	Lautwert	Zeichen	Lautwert	Zeichen	Lautwert	Zeichen	Lautwert	Zeichen	Lautwert
1 𐬀	a	10 𐬊	ā (nasal)	19 𐬑	č	28 𐬛	f	37 𐬤	v	46 𐬭	h
2 𐬁	ā	11 𐬋	i	20 𐬌	ǵ	29 𐬜	w	38 𐬥	v	47 𐬮	h'
3 𐬂	e	12 𐬔	ē	21 𐬕	t	30 𐬖	η	39 𐬦	r	48 𐬯	h̥v
4 𐬃	ē	13 𐬕	u	22 𐬖	d	31 𐬗	h̥	40 𐬧	s	49 𐬰	y
5 𐬄	ə	14 𐬖	ū	23 𐬗	ḃ	32 𐬘	n	41 𐬨	z	50 𐬱	št
6 𐬅	â	15 𐬗	k	24 𐬘	ḃ	33 𐬙	n, m	42 𐬩	š	51 𐬲	šc
7 𐬆	o	16 𐬘	g	25 𐬙	t	34 𐬚	m	43 𐬪	š	52 𐬳	ša
8 𐬇	ō	17 𐬙	h̥	26 𐬚	p	35 𐬛	y	44 𐬫	š	53 𐬴	
9 𐬈	â	18 𐬚	γ	27 𐬛	b	36 𐬜	y	45 𐬬	ž	54 𐬵	

Figure 8. Avestan alphabet from Haarmann 1990, showing 𐬥 *ii* and 𐬬 *uu*. Also shown are the ligatures *ša* 𐬱, *šc* 𐬲, and *št* 𐬳.





（コペンハーゲン大学図書館蔵 K1 (330 葉表) – Vidēvāt 21.2~3）

**Figure 9.** Manuscript of the Avestan *Vidēvāt* text from Kōno *et al.* 2001.

The blue text is in Avestan script with the ligatures *št*, *ah*, and *ša* marked in orange.

The remaining text in black is Pahlavi script, with Aramaic and Middle Persian spellings.

An error, scratched-out by the scribe, is shown in red in line 13.

1. [baēuu]arə.vāras.ciṭ mruīiā ašāum za-
2. raθuštra .: MDM ZK mtlg MYA l-LK w'l'n'
3. AMT slyškyh 'wptyt'W l-LK-c 'wptyt bywl w'l'n
4. -c YMRRRRWN'-m. z 'hlwb' zltwšt .: yaskahe a-
5. panaštahe mahrkahe apanašt-
6. ahe .: ysk BRA OZLWNšnyh l'y mlg BRA OZLWN-
7. šnyh l'y .: jaene-yaskahe apana-
8. štah jaene.mahrkahe apa. .: znyt'
9. ysk PWN BRA OZLWNšnyh MHYTWNyt nlg PWN BRA
10. OZLWNšnyh gaḍahe apa.gaḍahe
11. apa.naštahe .: ZK g'd' LWTE g'd' BRA
12. OZLWNšnyh 'pz'l Y OLEš'n .: yezi uzirō.-huua
13. mərəncaiti arəzahuua baēišaziiā-
14. t .: HT PWN 'pz'l mlncynyt PWN 'pyck wltšnyh
15. nyš'zynyt .: yezi arəzahuua mərəncaiti
16. xšapō.huua bišaziiāt .: HT PWN

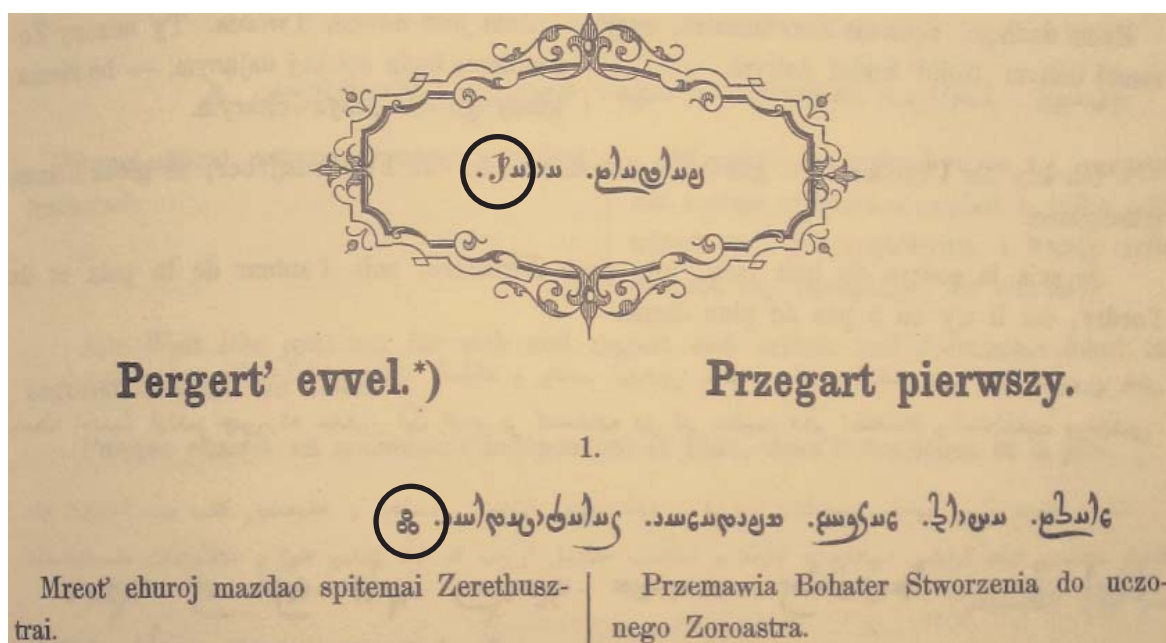
[illegible]

3

**Figure 10.** Printed text of the Avestan *Vidēvāt* text given in Figure 9 from Geldner 2003.

The blue text is in Avestan script with the ligatures 𐬨𐬀 *št* marked in orange. Compare this with Figure 9, where the ligatures 𐬨𐬀 *ah* and 𐬨𐬀 *ša* are used; these ligatures are not used in Geldner’s edition, because Geldner does not prefer the 𐬨𐬀 *ah* ligature, and because in this text he used 𐬨𐬀 *ša* not 𐬨𐬀 *ša* so the second ligature does not apply. These are shown in purple so they can be compared with Figure 9.

The Pahlavi text from the original is not represented.



**Figure 11.** Example from Pietraszewski 1858 showing AVESTAN LETTER LE.

The TURNED END OF SECTION is also represented



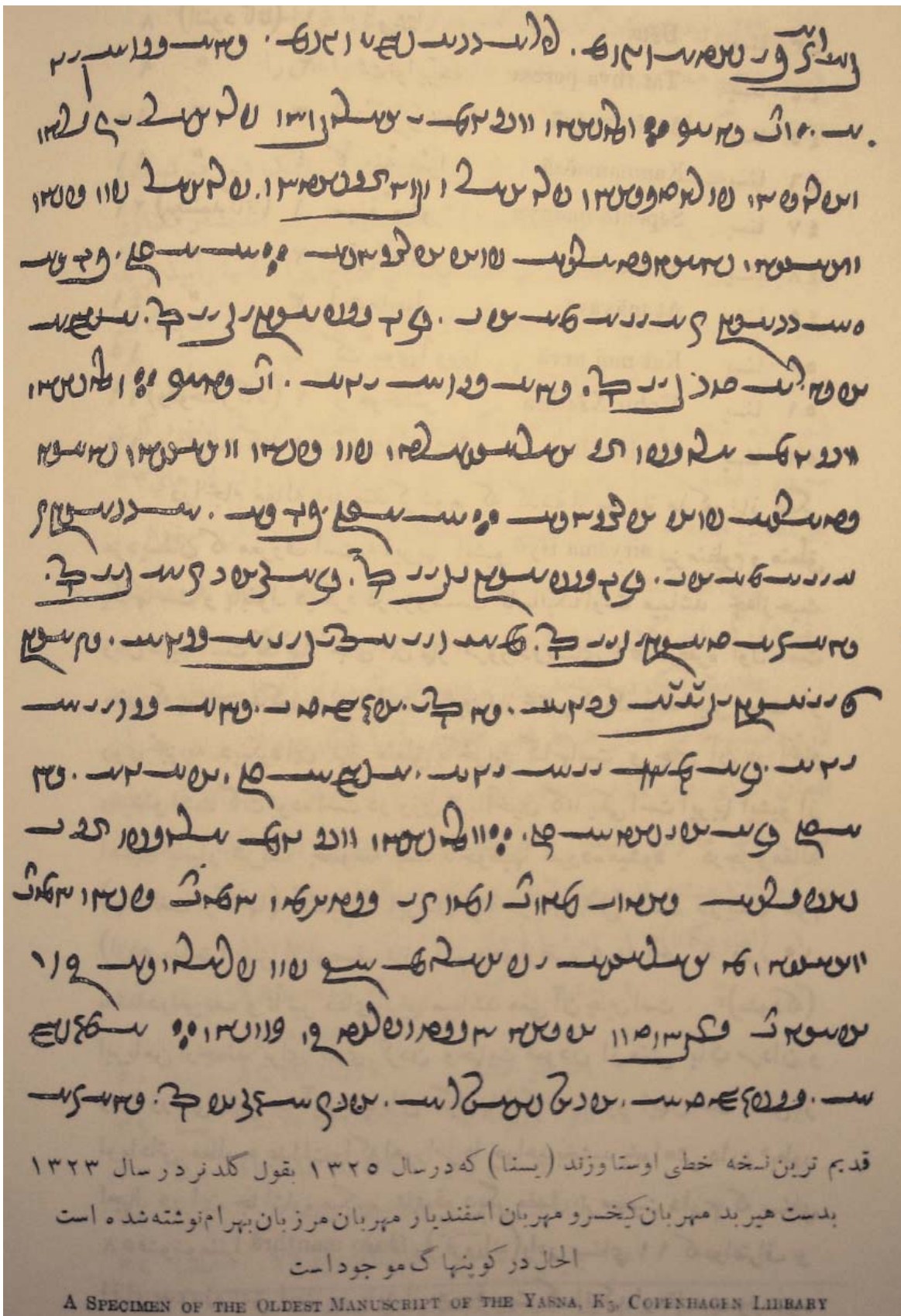


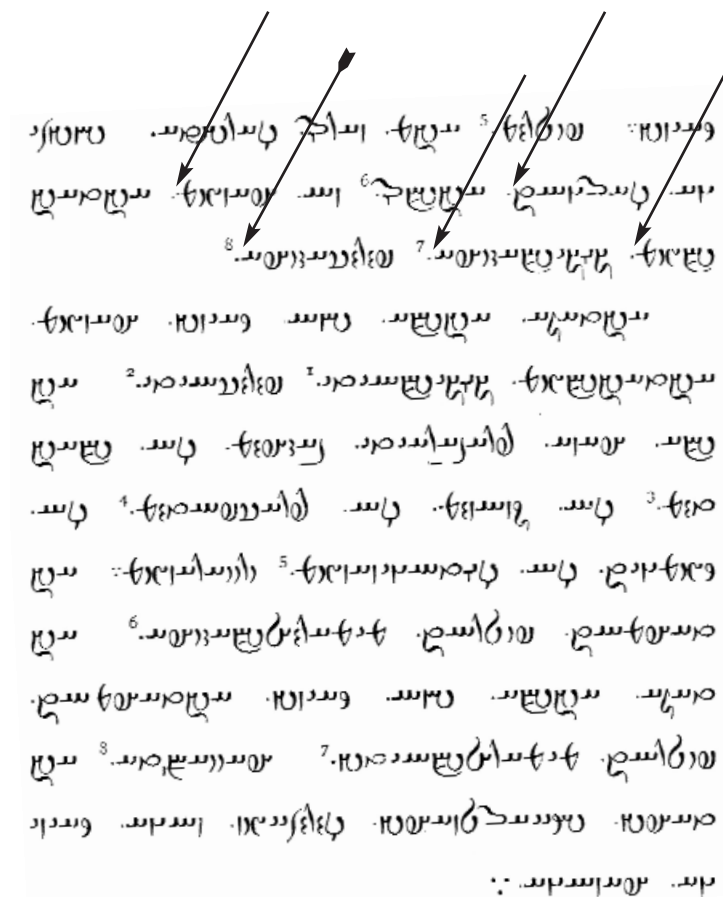
Figure 12. Sample from one of the earliest Yasna manuscripts, from Poure Davoud 1927.

H. Reichelt, Awestisches Elementarbuch. Heidelberg: Winter 1909, aber ohne genügende Berücksichtigung der Schriftgeschichte.

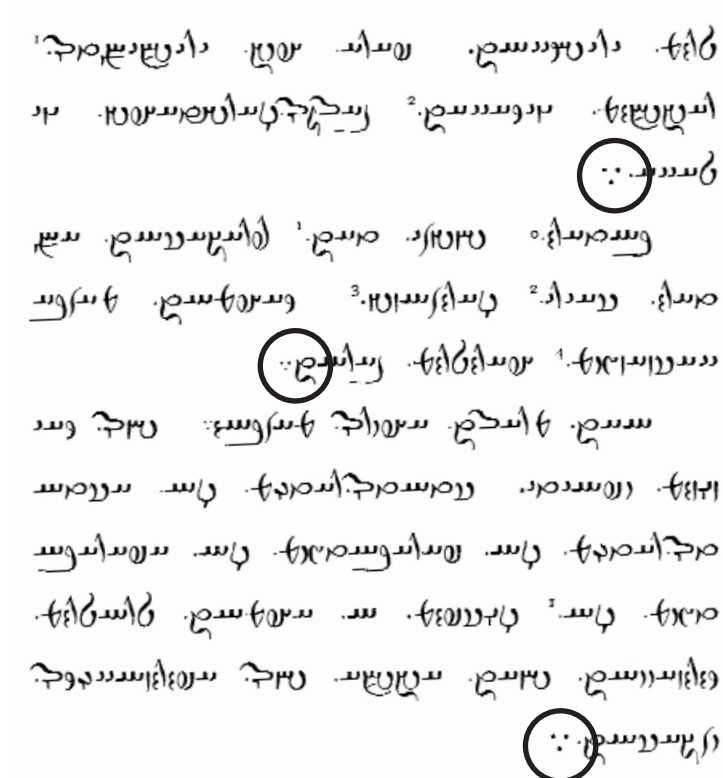
Die Awesta-Typen sind Eigentum der Göttinger Gesellschaft der Wissenschaften und von der Universitätsdruckerei Kastner in Göttingen zur Verfügung gestellt worden.

12





**Figure 15.** Passage of Avestan text from Geldner 1880 showing FULL STOP (with fletched arrow) alongside AVESTAN SEPARATION POINTS.

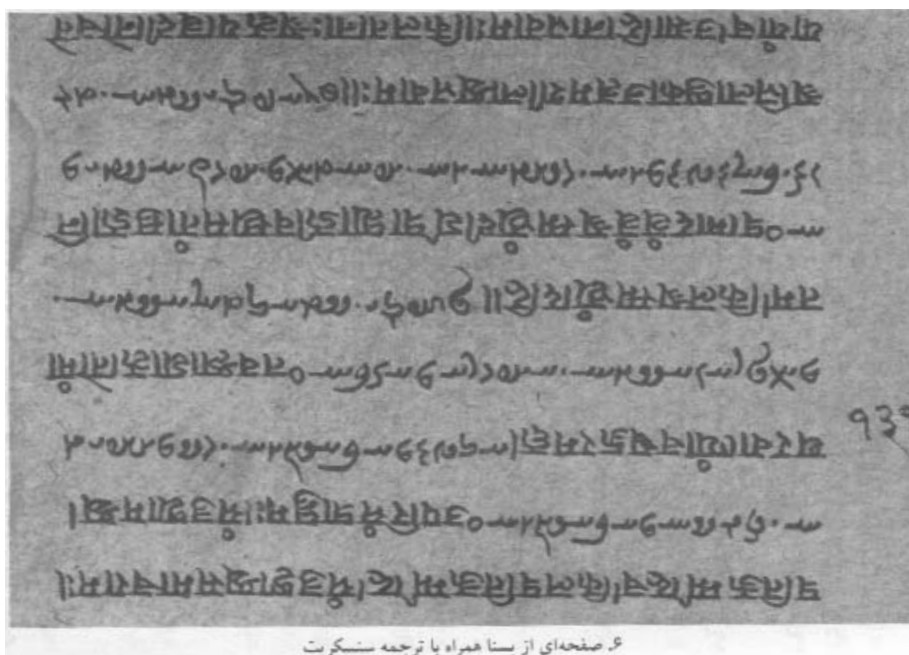


**Figure 16.** Passage of Avestan text from Geldner 1880 showing AVESTAN COLON (centre) alongside AVESTAN END OF SENTENCE (top and bottom).



Punctuation is a weak point in the manuscripts. They have various signs of punctuation, indeed, but they have no system of punctuation. I had to devise a system for myself, based upon the signs that are found: thus, : is a colon, ; semicolon or period, ∴ denotes the close of a paragraph or the end of a sentence, § a greater division, §§ the end of a chapter. In texts with Pahlavi and Sanskrit translation, this version takes the place of punctuation. The text is cut up by the translation into sentences of varied length, sometimes of only one or two words. For the most part, however, the explanation is inserted at an actual break in the sentence. It is only in the rarest instances that the manuscripts vary in this traditional division of the sentence.<sup>1</sup> In the sâda-texts, on the contrary, the text runs along without break until the first large division, with a mark of punctuation thrown in only here and there.<sup>2</sup> A general exception is made in the Gâthâs, where the close of the strophe is regularly denoted by the star. Where the translation was wanting I was therefore often thrown upon my own resources as regards punctuation. In that case the punctuation may be regarded in doubtful instances simply as subjective. In the Vendidad I have adhered almost absolutely to the division of sentences by the Pahlavi translation, and have indicated the same either by a larger point (.), or by ∴ where the sentence ends. I regret that I did not do this consistently also in the Yasna; although even there the larger point for the most part denotes a division by the translation.<sup>3</sup> Sometimes, however, I have departed from

**Figure 17.** Passage from Geldner's 1880 grammar discussing his systematic use of ∴ AVESTAN COLON, ∴ AVESTAN SEMICOLON, ∴ AVESTAN END OF SENTENCE, and § AVESTAN END OF SECTION, as well as his use of sentence-final FULL STOP, to be distinguished from AVESTAN SEPARATION POINT.



**Figure 18.** Passage of Avestan and Sanskrit text from Rashed Mohassel 2003. The text is interesting, because either the Avestan or the Sanskrit is inverted so that both have the same directionality. The Avestan text uses ligature 𐬨𐬀 *št* and ligature 𐬨𐬀 *ah*.

TABLE XXX - Row 08: AVESTAN

	080	081	082	083
0	𐬀	𐬁	𐬂	𐬃
1	𐬄	𐬅	𐬆	𐬇
2	𐬈	𐬉	𐬊	𐬋
3	𐬌	𐬍	𐬎	𐬏
4	𐬐	𐬑	𐬒	𐬓
5	𐬔	𐬕	𐬖	𐬗
6	𐬘	𐬙	𐬚	
7	𐬛	𐬜	𐬝	
8	𐬞	𐬟	𐬠	◦
9	𐬡	𐬢	𐬣	·
A	𐬤	𐬥	𐬦	∴
B	𐬧	𐬨	𐬩	∴
C	𐬪	𐬫	𐬬	∴
D	𐬭	𐬮	𐬯	∴
E	𐬰	𐬱	𐬲	◦◦
F	𐬳	𐬴	𐬵	◦◦

G = 00  
P = 00

TABLE XXX - Row 08: AVESTAN

hex	Name	hex	Name
00	AVESTAN LETTER A		
01	AVESTAN LETTER AA		
02	AVESTAN LETTER AO		
03	AVESTAN LETTER AAO		
04	AVESTAN LETTER AN		
05	AVESTAN LETTER AAN		
06	AVESTAN LETTER AE		
07	AVESTAN LETTER AEE		
08	AVESTAN LETTER E		
09	AVESTAN LETTER EE		
0A	AVESTAN LETTER O		
0B	AVESTAN LETTER OO		
0C	AVESTAN LETTER I		
0D	AVESTAN LETTER II		
0E	AVESTAN LETTER U		
0F	AVESTAN LETTER UU		
10	AVESTAN LETTER KE		
11	AVESTAN LETTER XE		
12	AVESTAN LETTER XYE		
13	AVESTAN LETTER XVE		
14	AVESTAN LETTER GE		
15	AVESTAN LETTER GGE		
16	AVESTAN LETTER GHE		
17	AVESTAN LETTER CE		
18	AVESTAN LETTER JE		
19	AVESTAN LETTER TE		
1A	AVESTAN LETTER THE		
1B	AVESTAN LETTER DE		
1C	AVESTAN LETTER DHE		
1D	AVESTAN LETTER TTE		
1E	AVESTAN LETTER PE		
1F	AVESTAN LETTER FE		
20	AVESTAN LETTER BE		
21	AVESTAN LETTER BHE		
22	AVESTAN LETTER NGE		
23	AVESTAN LETTER NGYE		
24	AVESTAN LETTER NGVE		
25	AVESTAN LETTER NE		
26	AVESTAN LETTER NYE		
27	AVESTAN LETTER NNE		
28	AVESTAN LETTER ME		
29	AVESTAN LETTER HME		
2A	AVESTAN LETTER YYE		
2B	AVESTAN LETTER YE		
2C	AVESTAN LETTER VE		
2D	AVESTAN LETTER RE		
2E	AVESTAN LETTER LE		
2F	AVESTAN LETTER SE		
30	AVESTAN LETTER ZE		
31	AVESTAN LETTER SHE		
32	AVESTAN LETTER ZHE		
33	AVESTAN LETTER SHYE		
34	AVESTAN LETTER SSHE		
35	AVESTAN LETTER HE		
36	(This position shall not be used)		
37	(This position shall not be used)		
38	AVESTAN ABBREVIATION MARK		
39	AVESTAN SEPARATION POINT		
3A	AVESTAN COLON		
3B	AVESTAN SEMICOLON		
3C	AVESTAN END OF SENTENCE		
3D	AVESTAN TURNED END OF SENTENCE		
3E	AVESTAN END OF SECTION		
3F	AVESTAN TURNED END OF SECTION		

## A. Administrative

1. Title

**Proposal to encode the Avestan script in the BMP of the UCS**

2. Requester's name

**UC Berkeley Script Encoding Initiative (Universal Scripts Project)**

3. Requester type (Member body/Liaison/Individual contribution)

**Liaison contribution.**

4. Submission date

**2006-10-20**

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

**Yes.**

6b. More information will be provided later

**No.**

## B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

**Yes.**

1b. Proposed name of script

**Avestan.**

1c. The proposal is for addition of character(s) to an existing block

**No.**

1d. Name of the existing block

2. Number of characters in proposal

**62.**

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

**Category C.**

4a. Is a repertoire including character names provided?

**Yes.**

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

**Yes.**

4c. Are the character shapes attached in a legible form suitable for review?

**Yes.**

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

**Michael Everson.**

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

**Michael Everson, Fontographer.**

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

**Yes.**

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

**Yes.**

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

**Yes.**

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

**See above.**

## C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

**Yes. See N2556, N1684.**

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

**Yes.**

2b. If YES, with whom?

**Hassan Rezai Baghbidi (Department of Linguistics, University of Tehran), Hossein Masoumi Hamedani (Iranian Academy of Persian Language and Literature), Ali Reza Mohazab, Jost Gippert (TITUS Projekt), Desmond Durkin-Meisterernst, Günter Schweiger**

2c. If YES, available relevant documents

**<http://titus.fkdig1.uni-frankfurt.de/unicode/iranian/3tagung.htm>**

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

**Zoroastrians, Iranianists and other scholars.**

4a. The context of use for the proposed characters (type of use; common or rare)

**Used liturgically and by scholars.**

4b. Reference

5a. Are the proposed characters in current use by the user community?

**Yes.**

5b. If YES, where?

**Religious and scholarly publications.**

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

**Yes.**

6b. If YES, is a rationale provided?

**Yes.**

6c. If YES, reference

**Accordance with the Roadmap. Avestan is used in modern Zoroastrian religion.**

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

**No.**

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

**No.**

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

**No.**

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

**Yes.**

10b. If YES, is a rationale for its inclusion provided?

**Yes.**

10c. If YES, reference

**U+0838 AVESTAN ABBREVIATION SIGN bears a superficial resemblance to other rings in the UCS, such as U+02DA RING ABOVE, or U+2218 RING OPERATOR, but those are true circles, with different origins and uses, having nothing to do with Avestan; example of the rings also show that they are typically not drawn as true circles, but with strokes appropriate to Avestan letterforms. U+0970 DEVANAGARI ABBREVIATION SIGN would be more appropriately analogous, but a unification of the two would be inconvenient to users. U+0839 AVESTAN SEPARATION POINT looks like U+002E FULL STOP but is smaller and used alongside it in Avestan; it also could be said to look like U+10101 AEGEAN WORD SEPARATOR DOT, but we consider this to be a far-fetched proposal. The small 3-dot punctuations U+083A through U+083D look superficially similar to U+2056 THREE DOT PUNCTUATION, but the orientation is different and significant. The three down-pointing dot configurations have a particular size relation to each other within the context of Avestan. Since Avestan and the as-yet unencoded Pahlavi both share this punctuation, we are convinced that this collection of punctuation marks should be encoded as script-specific, strong RTL punctuation. The recently added U+061E ARABIC TRIPLE DOT PUNCTUATION MARK cannot be unified with the Avestan and Pahlavi punctuation either. Here again, the three Avestan characters have a specific size relation to each other, and the Arabic punctuation character is both historically unrelated to them, and needs to be consistently represented in proportion to Arabic letters and other punctuation; it also has the typical Arabic diamond-shape to its dots.**

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

**No.**

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

**No.**

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

**No.**

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

**No.**

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?